Nucleic Acids Research, 1997, Vol. 25, No. 8 **1537–1547**

Structural features of the DNA hairpin d(ATCCTA-GTTA-TAGGAT): formation of a G-A base pair in the loop

Maria J. P. van Dongen, Margret M. W. Mooren, Esther F. A. Willems, Gijs A. van der Marel[1], Jacques H. van Boom[1], Sybren S. Wijmenga[§] and Cornelis W. Hilbers*

NSR Centre for Molecular Structure, Design and Synthesis, Laboratory of Biophysical Chemistry, University of Nijmegen, Nijmegen, The Netherlands and [1]Gorlaeus Laboratories, Department of Organic Chemistry, State University of Leiden, Leiden, The Netherlands

Received January 10, 1997; Revised and Accepted February 21, 1997

ABSTRACT

The three-dimensional structure of the hairpin formed by d(ATCCTA-GTTA-TAGGAT) has been determined by means of two-dimensional NMR studies, distance geometry and molecular dynamics calculations. The first and the last residues of the tetraloop of this hairpin form a sheared G-A base pair on top of the six Watson–Crick base pairs in the stem. The glycosidic torsion angles of the guanine and adenine residues in the G-A base pair reside in the *anti* and high-*anti* domain (~–60°) respectively. Several dihedral angles in the loop adopt non-standard values to accommodate this base pair. The first and second residue in the loop are stacked in a more or less normal helical fashion; the fourth loop residue also stacks upon the stem, while the third residue is directed away from the loop region. The loop structure can be classified as a so-called type-I loop, in which the bases at the 5′-end of the loop stack in a continuous fashion. In this situation, loop stability is unlikely to depend heavily on the nature of the unpaired bases in the loop. Moreover, the present study indicates that the influence of the polarity of a closing A·T pair is much less significant than that of a closing C·G base pair.

INTRODUCTION

For many years considerable effort has been put into conformational studies aimed at establishing the structural principles that underly DNA and RNA hairpin loop stability (reviewed in 1). From these investigations, which were mainly focused on hairpins with tetraloops, it has emerged that stable loops very often possess a base pair between the first and last loop residue. The amount of stabilization seems to depend on the exact nature of this base pair, which can be either canonical (C·G) or non-canonical. Recently, a particular group, i.e. loops with a G-A base pair, has gained much interest (2–9). Sequences which have the potential to form such a loop are generally remarkably stable and, moreover, fairly common. Tetraloops with the consensus sequence GNRA (in which N is any nucleotide and R is purine), for example, are abundant motifs in rRNA (5,10). DNA hairpins comprising G-A base pairs in the loop have been related to telomeric and centromeric structures (11,12).

An important part of the current knowledge on G-A base pairs has resulted from studies performed on (partly) self-complementary oligonucleotide strands, which form duplex structures with a tandem G-A mismatch (13–18). In particular, it has been found that structures with two successive, so-called sheared G-A base pairs formed by a 5′-GA-3′/3′-AG-5′ sequence are exceptionally stable. This stability has primarily been attributed to very favourable inter-strand GG and AA stacking interactions (15). According to heuristic rules, postulated on the basis of these studies, the formation of this stable structural unit is promoted if the guanine residue is preceded by a pyrimidine residue (19,20).

Remarkably, conformational studies on hairpins with the potential to form a G-A base pair in the loop have thus far revealed, in all instances, the existence of the same sheared G-A base pair configuration (4,5,7–9). Based on the results of these investigations it has therefore been suggested that favourable stacking interactions between the G-A and the underlying base pair in the stem, preferably C·G (8,21), contribute to the stability of these hairpins. In addition, RNA GNRA loops seem to be further stabilized by the formation of various hydrogen bonds involving phosphate oxygen and hydroxyl oxygen atoms (5,7,22).

The influence of the unpaired bases in the loop has thus far been a matter of debate. Experiments by Miura and co-workers (2,3) indicate that this base(s) does play a substantial role. For instance, they observed melting temperatures for hairpins formed by d(GC-GAAA-CG) and d(GC-GAA-CG) which are ~7°C higher than that of d(GC-GTTA-GC). NMR studies on the GAA loop support this view. Extensive stacking interactions were found between the base of the second loop residue and the G-A base pair formed by the first and third loop residues (4). Interesting results

*To whom correspondence should be addressed. Tel: +31 243 652 678; Fax: +31 243 652 112; Email: ceesh@sci.kun.nl

Present addresses: [+]Unilever Research Laboratories, Vlaardingen, The Netherlands and [§]Department of Medical Biochemistry and Biophysics, University of Umeå, Sweden

Table 1. NMR experiments performed for the d(ATCCTA-GTTA-TAGGAT) fragment

Experiment[a]	Mixing time (ms)	Frequency (MHz)[b]	Number of data points[b]	Spectral width (kHz)[b]	Solvent	Temperature (K)	Reference
jr-1D		400	8192	8065	H_2O	271, 276, 282, 288, 294	(43)
jr-NOESY	300	400	2048/496	8065/8065	H_2O	276	(43,44)
1D		400	16384	4000	D_2O	298	
NOESY	300	600	2048/512	6000/6000	D_2O	298	(44)
NOESY	75, 100, 150	400	2048/512	4000/4000	D_2O	298	(44)
DQF-COSY		600	4096/512	6000/6000	D_2O	298	(45,46)
TOCSY	100	400	2048/512	4000/4000	D_2O	298	(47,48)
^{31}P 1D		81	4096	1000	D_2O	298	
^{31}P-^{1}H correlation		400/162	2048/282	4000/400	D_2O	298	(49)

[a]For two-dimensional experiments phase-sensitive detection of the indirectly observed frequency was achieved by the TPPI method (45).
[b]If two numbers are given, the first one refers to the direct detected dimension (f2), while the second number refers to the indirect dimension (f1).

have also been obtained by Reid and co-workers, who report completely different conformations for d(CAAT-GCA-ATG) and d(GAAT-GGA-ATG). The first sequence forms a hairpin with a sheared G-A base pair within a three-membered loop, whereas the second forms a bimolecular duplex, involving a $(GGA)_2$ motif (9). At the same time, this precludes definitive conclusions about a possible role of the unpaired bases, because the hairpin has to compete with the dimeric form. Others, including our group, have not observed such a large effect of the nature of unpaired bases in the loop. Zuiderweg and collaborators compared the melting temperatures of a variety of hairpins containing the sequence d(C-GNNA-G) and found a negligible influence of the N bases on stability (8). Thermodynamic studies performed in our laboratory on d(ATCCTA-GAAA-TAGGAT) and d(ATCCTA-GTTA-TAGGAT) revealed that these hairpins are about equally stable, with melting temperatures of 53 and 55°C respectively (23).

Here we present the results of a detailed determination of the loop structure of d(ATCCTA-GTTA-TAGGAT) by means of NMR. This DNA fragment, which was designed to form a hairpin with a G-A base pair in the four-membered loop, differs from that in an earlier NMR study, d(TGC-GGCA-GCA) (8), in two essential aspects. First, in the hairpin studied in this paper the first unpaired residue in the loop is a pyrimidine (T) instead of a purine (G) residue and, second, the stem is closed by a purine·pyrimidine (A·T) instead of a pyrimidine·purine (C·G) base pair. The present study shows that the loop structure can be classified as a so-called type-I loop (24), in which the bases at the 5′-end of the loop stack in a more or less continuous fashion. In this situation, the stability of the loop is unlikely to depend heavily on the nature of the unpaired bases in the loop. In addition, the present study indicates that the polarity of a closing A·T base pair influences the loop stability much less significantly than does a C·G pair.

MATERIALS AND METHODS

NMR spectroscopy

The oligonucleotide d(ATCCTA-GTTA-TAGGAT) was synthesized by the phosphotriester method (25). NMR samples were prepared by dissolving three times lyophilized oligonucleotide to a concentration of 3 mM in a D_2O solution containing 25 mM sodium phosphate, 1 mM sodium cacodylate and 0.1 mM EDTA, pH 6.8. NaCl was added to a total sodium concentration of 200 mM. A similar sample in a 95% H_2O/5% D_2O solution was prepared. Parameters and conditions for the NMR experiments, recorded on Bruker AM-400 and AM-600 spectrometers interfaced to ASPECT 3000 computers, are summarized in Table 1. Data processing was carried out using either UX-NMR software running on a Bruker-X32 computer or the M-NMR software package running on a Silicon Graphics Indigo$^{2\text{TM}}$ work station.

Estimation of structure restraints

Volumes of NOE cross-peaks of the NOESY spectra with mixing times of 75, 100, 150 and 300 ms were integrated using UX-NMR and served as input for the NO2DI program (26), running on a CONVEX-C120 computer. The NO2DI algorithm calculates inter-proton distances from NOE intensities, taking spin diffusion into account. The calculations were performed with the rotation correlation time set to 2.0 ns, as estimated with the aid of the Stokes–Einstein relation (27). The NOEs belonging to the H5–H6 proton pairs of cytosines were used for calibration. Error bounds of –10 and +20% of the calculated distances were assumed to produce lower and upper distance bounds respectively.

$J_{H1'H2'}$ and $J_{H1'H2''}$ coupling constants were derived by simulation of the one-dimensional proton spectrum, using PANIC (Bruker). Sums of J-couplings, $\Sigma 2'$ and $\Sigma 2''$, were estimated from multiplet patterns in a DQF-COSY spectrum. For $J_{1'2'}$ and $J_{1'2''}$, errors of 0.3 Hz were assumed, based on the digital resolution of the NMR spectrum. For $\Sigma 2'$ and $\Sigma 2''$, the errors were estimated as 5.0 Hz for residues C3 and G13, 3.0 Hz for T2, G7, A12 and G14 and 1.5 Hz for the remaining residues. The allowed pseudorotation parameters, P and ϕ_m, of the sugars were derived from these J-couplings using the program MARC (28,29). P_S was sampled between 117 and 211.5° in steps of 4.5°, ϕ_{mS} was sampled between 32 and 42° in steps of 2° and the molar fraction f_S was sampled between 0.5 and 1.0 in steps of 0.005. P_N and ϕ_{mN} were kept fixed at 9 and 35° respectively.

Table 2. X-PLOR protocols used to calculate the conformation of the d(ATCCTA-GTTA-TAGGAT) fragment

	Simulated annealing			Refinement		Energy minimisation
	Conformational Search	Cooling	Minimisation	Cooling	Minimisation	
Temperature (K)	1000	1000 → 300	–	1000 → 300	–	–
Energy constants	1000	1000	1000	1000	1000	1000
K_{bond} (kcal·mol^{-1}·Å^{2})	500	500	500	500	500	500
K_{angle} (kcal·mol^{-1}·rad^{-2})[b]	–	–	–	–	–	–
K_{impr} (kcal·mol^{-1}·rad^{-2})[b]	500	500	500	500	500	500
K_{planar} (kcal·mol^{-1}·rad^{-2})	10	10	10	10	10	–
K_{repel} (kcal·mol^{-1}·Å^{-4})[c]	0.003	0.003 → 4	4	0.003 → 4	4	–
K_{LJ} (kcal·mol^{-1}·Å^{-4})	–	–	–	–	–	1
K_{elec} (kcal·mol^{-1}·Å^{-2})	–	–	–	–	–	1
K_{NOE} (kcal·mol^{-1}·Å^{-2})[d]	50	50	50	50	50	50
K_{cdih} (kcal·mol^{-1}·rad^{-2})	5	200	200	200	200	200
Length of simulation	18 ps	9 ps	200 steps	9 ps	200 steps	2000 steps

[a]K_{bond}, K_{angle}, K_{impr}, K_{planar}, K_{repel}, K_{LJ}, K_{elec}, K_{NOE}, K_{cdih} stand for the force constants associated with the energy terms for bond lengths, bond angles, 'improper' dihedral angles, planarity restraints, non-bonded repulsion, Lennard–Jones interaction, electrostatic interaction, distance restraints and dihedral restraints respectively.

[b]During the first 12 ps of the conformational search the weighting factor for the bond and 'improper' energy were reduced by multiplication by a factor 0.4 and 0.1 respectively.

[c]The force constant for the repel function was coupled to a multiplication factor for the van der Waals radii, i.e. for a force constant of 0.003 kcal·mol^{-1}·Å^{-4} the radii were multiplied by a factor 0.9; for a force constant of 4 kcal·mol^{-1}·Å^{-4} this factor was 0.75.

An harmonic flat-bottom potential with linear behaviour for deviations >0.5 Å was used. During the first 12 ps of the conformational search the slope of the asymptote was set to 0.1; in later stages this slope was set to 1.

Ranges allowed for the backbone dihedral angles β, γ and ε as well as the glycosidic torsion angle χ were derived from NOE data combined with homo- and heteronuclear J-coupling constants (*vide infra*).

Generation of starting structures

The thus derived distance and dihedral angle ranges served as input for the variable target function algorithm DIANA, version 2.8 (30,31), adapted in our laboratory to incorporate variable sugar puckers. Calculations were performed on a Silicon Graphics Indigo^{2}™ workstation. For the stem of the hairpin standard B-DNA distance restraints, hydrogen bond restraints and dihedral angle ranges (29,32,33) were used, as well as restraints to preserve base pair planarity.

The DIANA structures with the lowest target function were used as input for the program CORMA, version 2.25 (34), which calculates a dipole–dipole relaxation matrix for a system of protons and converts that into intensities expected for a NOESY experiment. A number of cycles of DIANA and CORMA calculations were performed in the same way as described before (24,35), to obtain a model for the hairpin with a minimum number of violations of the NMR data.

The final DIANA input included 72 upper and lower distance restraints. An additional set of 65 lower bound distance restraints of 4.0 Å was added for the loop region. They followed from the first cycles of DIANA calculations, which generated structures with short inter-proton distances for which no corresponding experimental NOE cross-peak was observed while at the same time other NOEs involving the same protons ruled out local motion. A total of 50 lower and upper distance restraints were used to preserve the base pairs, including four lower and upper distance restraints for the H bonds of a G-A base pair. Furthermore, 98 dihedral restraints in the stem and stereospecific assignments for 25 out of 32 CH_2 groups (H2′/H2″ or H5′/H5″) were used in the calculations.

Two separate DIANA calculations, representing two distinct, possible ranges for the β dihedral angle of residue 9 (*vide infra*), were performed, each producing a total of 200 structures. For each conformation the target function was minimized at the so-called levels 0–15 (see 31), whereby minimization at level 15 was done three times. The weighting factors for explicit upper and lower distance limits were set to 1.0 and the weighting factor for dihedral angle restraints was set to 5.0. For levels 0–14 and for the first minimization step at level 15 the weighting factor for the van der Waals restraints was 0.2, for the second minimization step at level 15 it was increased to 0.5 and for the final minimization step at level 15 to 1.0. For minimization levels 0–14 the maximal number of target function calculations was set to 100; for level 15 it was set to 200.

Structure refinement

The 15 structures with smallest target function values (<3.0 Å^{2}) of each set were used as input for X-PLOR, version 3.1 (36), which was run on a Silicon Graphics Indigo^{2}™ workstation. They were subjected to simulated annealing and refinement protocols, not substantially different from those delivered with the X-PLOR software, of which the details are given in Table 2. For each input structure five new structures were generated by random variation

Table 3. Chemical shifts of d(ATCCTA-GTTA-TAGGAT) at 298K[a]

residue	NH[b]	NH_2[b]		H6 H8	H5 CH_3 H2	H1'	H2'	H2''	H3'	H4'	H5'	H5''	P
A1	–			8.25	7.99	6.23	2.75	2.90	4.87	4.28			–
T2	13.85	–	–	7.43	1.38	6.06	2.27	2.55	4.90	4.31	4.20	4.15	-3.74
C3	–	8.29	6.82	7.62	5.65	6.06	2.32	2.56	4.85	4.28	4.20	4.16	-3.60
C4	–	8.40	6.98	7.57	5.60	5.90	2.10	2.50	4.80	4.19			-3.49
T5	13.73	–	–	7.24	1.72	5.70	**1.65**	2.14	4.81	4.07			-3.77
A6	–	7.73	6.73	7.87	7.58[b]	6.13	2.32	2.48	4.93	4.39	4.06	4.01	-3.80
G7	10.89			8.04	–	5.57	**2.72**	**2.46**	4.95	4.49	4.25	4.14	-4.27
T8	10.89	–	–	7.26	1.56	5.71	**1.91**	**2.05**	4.47	**3.35**	3.73[c]	3.75[c]	-3.48
T9	10.89	–	–	7.37	1.64	6.03	2.10	2.33	4.60	**3.80**	**3.37**[c]	3.75[c]	-3.92
A10	–			8.09	7.97	6.24	2.67	2.80	4.82	4.34	3.99	3.92	-3.72
T11	13.46	–	–	7.30	1.90	**5.07**	1.89	2.26	4.77	4.11		4.26	-4.19
A12	–	7.96	6.50	8.17	7.58[b]	5.96	2.73	2.87	5.04	4.40	4.12	4.01	-3.55
G13	12.80			7.62	–	5.54	2.54	2.66	5.06	4.36	4.17	4.18	-3.60
G14	12.64			7.60	–	5.67	2.53	2.71	4.96	4.36		4.17	-3.39
A15	–	7.66	6.17	8.12	7.99	6.26	2.62	2.88	4.97	4.43			-3.54
T16	13.31	–	–	7.20	1.44	6.12	2.14	2.17	4.51				-3.66

[a]Proton chemical shifts relative to DSS; phosphorus chemical shifts relative to TMP. Unusual values are shown in bold.
[b]Chemical shift obtained from NOESY spectrum recorded for the H_2O sample at 276K.
[c]Ambiguous stereo-specific assignment.

Table 4. J-coupling constants and sugar pucker parameters of d(ATCCTA-GTTA-TAGGAT)

residue	$J_{1'2'}$	$J_{1'2''}$	$\Sigma J_{H2'}$	$\Sigma J_{H2''}$	P	ϕ_m	f_S
A1	–	–	25	23	198–212	36–42	0.75–0.90
T2	8.5	5.5	25	23	158–198	40–42	0.75–1.00
C3	8.8	5.8	24	22	135–203	35–42	0.75–1.00
C4	7.1	6.9	29	23	139–212	32–34	0.70–1.00
T5	7.0	7.0	27	23	185–212	32–34	0.75–0.95
A6	6.9	6.9	26	22	198–212	32–40	0.80–0.90
G7	9.3	5.1	25	21	171–189	42	0.85–1.00
T8	–	–	27	23	126–212	32–42	0.65–1.00
T9	7.3	6.9	28	22	185–212	32–34	0.80–1.00
A10	–	–	26	23	153–212	32–42	0.65–1.00
T11	7.9	6.7	29	21	185–203	32	0.90–1.00
A12	9.5	6.0	27	21	153–189	32–38	0.90–1.00
G13	10.6	4.8	25	21	149–176	40–42	0.95–1.00
G14	–	–	24	21	157–212	32–42	0.65–1.00
A15	–	–	28	23	117–212	32–42	0.60–1.00
T16	7.2	7.2	–	–	117–126	32–34	0.65–0.70

of the initial velocities, resulting in 75 structures in total. As in the DIANA calculations, two possible ranges for β(9) were distinguished. The force field used for these calculations consisted of geometric terms accounting for bond stretching, bond angle bending and maintaining planar ring structures and (pro)chiral centres. The standard parameter and topology files were supplemented by 'improper' terms to maintain prochirality around the C5′ atoms. Performing the calculations *in vacuo*, a soft repulsive term was used for the non-bonded interactions. The NMR-derived distance and dihedral restraints were added as (soft) square well potentials. The base pairs of the stem were kept flat using weak planarity restraints. After these calculations a structure was accepted as a reasonable conformation if the sum of energies resulting from NMR-derived distance and dihedral restraints did not exceed 75 kcal/mol.

As a last step, the accepted structures were subjected to restrained energy minimization in a force field including terms for electrostatic and dihedral contributions. The effect of atomic charges was reduced by the application of partial charges and by choosing a distance-dependent dielectric constant, $\varepsilon = \varepsilon_0 \cdot r$. During this stage, the soft repulsive non-bonded energy term was replaced by a Lennard–Jones potential; the distance restraints used earlier to preserve the base pairs of the stem and the H bonds of the G-A base pair were omitted, as were the planarity restraints.

Several specific interaction energy contributions were calculated using the V_{pert} function in X-PLOR. Quanta (Polygen), running on a Silicon Graphics Indigo^{2TM} workstation, was used to visualize the results of the structure calculations. Helical parameters were calculated with the program NEWHELIX93 (kindly provided by R.E.Dickerson).

RESULTS AND DISCUSSION

Assignments

The imino proton spectrum of the hairpin with GTTA loop, recorded at several temperatures, is shown in Figure 1. The assignments of the imino proton and amino proton resonances are summarized in Table 3. They were derived from NOESY spectra recorded for the sample dissolved in H_2O (not shown). Between 12.0 and 15.0 p.p.m. the resonances originating from base paired residues are detected and account for all base pairs in the stem.

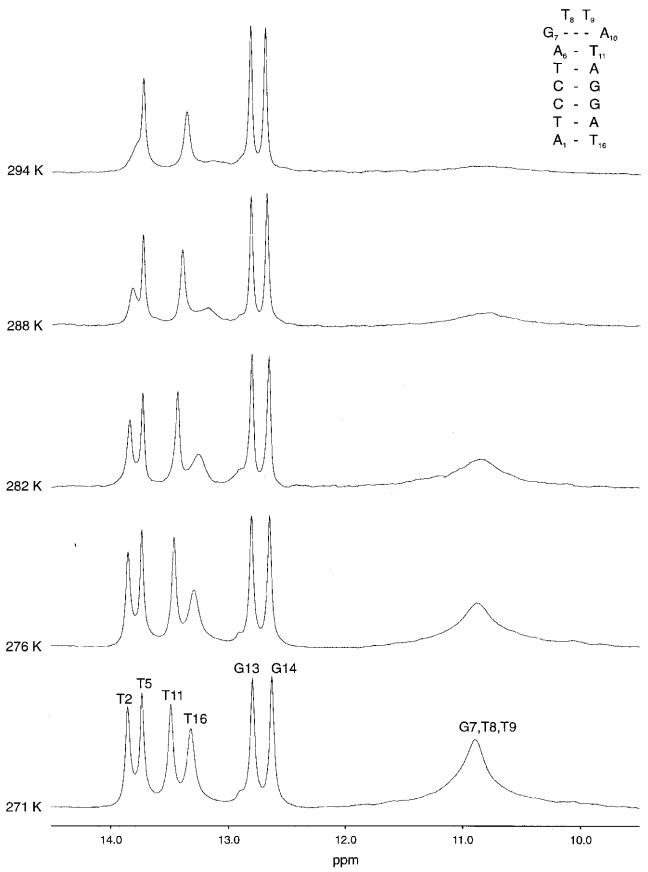

Figure 1. Imino spectra of d(ATCCTA-GTTA-TAGGAT), recorded as a function of temperature.

The integrated peak intensity in the spectrum near 10.9 p.p.m. accounts for three resonances. The imino proton resonances of the unpaired T8 and T9 are observed at this position and the imino proton of G7 accounts for the intensity of the third resonance. The resonance position and temperature-dependent behaviour of the imino proton of G7 indicate that this proton is not involved in hydrogen bonding.

Spectra of the hexadecanucleotide d(ATCCTA-GTTA-TAGGAT) dissolved in D_2O were recorded at room temperature. As an example, Figure 2a presents a spectral region of the NOESY spectrum, recorded with a mixing time of 300 ms, in which base–H8/H6 and the sugar–H1′ cross-peaks are observed. The proton resonances visible in these spectra were assigned using a standard sequential analysis procedure (29), the results of which are summarized in Table 3. It is noted that for the majority of residues a stereospecific assignment for the H5′ and H5″ spins could be made.

A few unusual upfield shifts of proton resonances were noticed. The resonance positions of H4′ and H5′ protons of T9, for example, show considerable upfield shifts. This generally indicates that these resonances are influenced by ring currents of a base plane positioned directly above or under the protons. Furthermore, the position of the H4′ resonance of T8 is more upfield than those of H5′ and H5″ resonances. Again, this effect suggests the presence of a base plane directly above or below this proton. Also worth noting is the observation that the relative chemical shifts of the H2′ and H2″ resonances of G7 are reversed compared with the common situation. The H2″ proton of this residue resonates more upfield than the H2′ proton, an effect that has sometimes been related to the presence of a sheared G-A base pair configuration (15,16). Finally, the chemical shift observed for the H1′ resonance of T11 (5.07 p.p.m.) is unusual. Similar upfield shifts have been observed for the H1′ resonances in other fragments with a G-A base pair in the loop. In those fragments, unusual ring current effects of the adenine base, due to a substantially reduced helical twist of the G-A pair, account for this observation (5,7).

Subsequently, the ^{31}P spectrum of d(ATCCTA-GTTA-TAGGAT) was investigated by means of a ^{1}H-^{31}P heterocorrelation experiment (Fig. 3). Using the earlier derived proton chemical shifts, almost all phosphorous resonances could be assigned (Table 3). It is noticeable that the resonances of P(G7) and P(T11) show an upfield shift of ~0.5 p.p.m. relative to the values for the other residues. None of the phosphorous resonances shows a substantial downfield shift, which would indicate a B_{II}-phosphate conformation (*–gauche*, *+trans*) (15).

NOEs in the loop region

Most sequential NOEs expected for helical regions, e.g. between H6/H8 and H1′ resonances (Fig. 2a) and between H6/H8 and H2′/H2″ resonances, were also visible for the GTTA loop region. Although the H1′(T8)–H6(T9), H2′/H2″(T8)–H6(T9) and H2″(T9)–H8(A10) cross-peak intensities are unusually weak, only the H1′(T9)–H8(A10) and H2′(T9)–H8(A10) NOEs could not be observed. Figure 2b summarizes these results.

In addition to the standard sequential sugar–base connectivities, the NOESY spectra showed a large number of unusual contacts between the various loop residues. Figure 4 schematizes the full set of inter-residue NOEs observed for the GTTA loop region. The connectivities found for the A6–G7, G7–T8 and A10–T11 steps indicate that the first and bases of the loop, G7 and A10 respectively, are stacked upon the terminal base pair of the stem region, A6-T11, and that residue T8 is more or less stacked upon residue G7. The lack of intense NOEs to T9 suggests that this residue must be directed away from the loop.

Determination of sugar conformations and backbone dihedral angles

A J-coupling analysis of the resonances of the sugar moieties was performed and the results were used to determine the sugar conformations. The results are listed in Table 4. The sugar rings are in an S-type conformation, although for almost all residues N-type conformations are admixed to some extent. It should be noted that the J-coupling information for residues A1, T8, A10, G14, A15 and T16 is rather limited. The sugar conformations of these residues are consequently not well defined (see Table 4). A qualitative analysis of H3′–H4′ cross-peak intensities in the TOCSY spectrum, which depend on the values of the $J_{H3'H4''}$ coupling constants, corroborated the results. For A6, G7, A10, T11 and A12 these cross-peaks are either absent or show very low intensities. This suggests that the sugar rings of these residues are present almost exclusively in the S-type form. For the remaining residues, among which are included T8 and T9, the TOCSY data indicate some conformational averaging.

The ^{1}H-^{31}P heterocorrelation spectrum (Fig. 3) provided information on the β angles (P-O5′-C5′-C4′). The cross-peaks

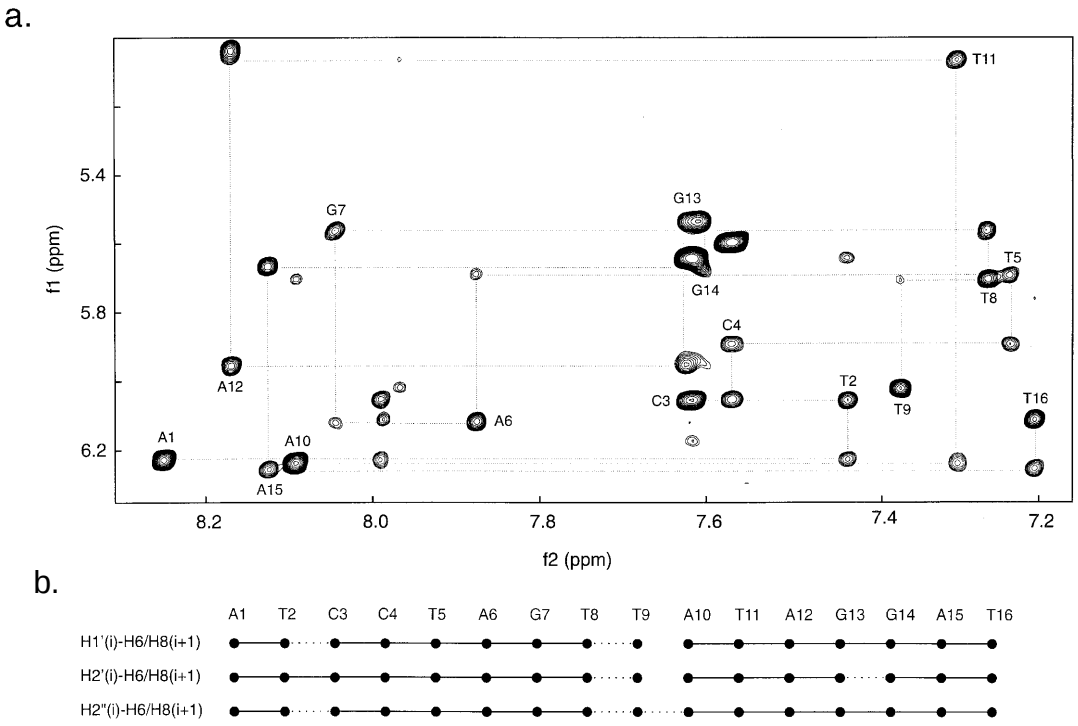

Figure 2. (**a**) Spectral region of the NOESY spectrum (mixing time 300 ms) of d(ATCCTA-GTTA-TAGGAT) at 298 K, containing H6/H8–H1′/H5 connectivities. Lines and labels indicate the sequential walk. (**b**) Connectivity diagram in which the presence of NOEs between neighbouring residues is indicated by solid lines. A blank region corresponds to the absence of an NOE. Very weak cross-peaks and cross-peaks which suffer from serious overlap are indicated by dashed lines.

observed for P(G7), i.e. H3′(A6)–P(G7), H4′(G7)–P(G7) and H5′(G7)–P(G7), imply that the β angle of G7 is in the *trans* conformation, ~150°. P(T8) shows cross-peaks to H3′(G7), H4′(T8) and H5″(T8), as in standard B-type DNA, showing that β(8) adopts a value of ~210°. For P(T9) the cross-peaks to H5′(T9) and H5″(T9) are both present, while the cross-peak to H4′(T9) is absent. From cross-sections through the multiplets it was estimated that $J_{H5'P}$ as well as $J_{H5''P}$ are ~7–10 Hz. This indicates that β for T9 adopts a value of ~110 or ~250°, dependent upon the stereospecific resonance assignment of H5′/H5″ of this residue. P(A10) shows cross-peaks to H3′(T9) and H5″(A10). It was estimated from the cross-section through the A10(H5″) multiplet that $J_{H5''P}$ is ~7–10 Hz. The standard cross-peak H4′(A10)–P(A10) is absent, which implies that β may still adopt a value in the *trans* region of ~210°, but γ must deviate from the standard *gauche*(+) value (*vide infra*).

The γ angles (O5′-C5′-C4′-C3′) were determined by the combined use of H3′/H4′–H5′/H5″ NOE intensities and estimations of J-coupling constants $J_{4'5'}$ and $J_{4'5''}$ from a DQF-COSY spectrum (28). It was deduced that γ of residue A6 is ~90° and γ of G7 is ~60°. γ of T8 could not be established unambiguously in this way because of overlap. The presence of a P(8)–H4′(8) cross-peak in the ^{1}H-^{31}P heterocorrelation, however, demonstrated that the value must be ~60°. The γ angle of T9 was found in the *trans* region, while the γ angle of A10 was deduced to be in the range 100–150°.

Information on the ε (C4′-C3′-O3′-P) angles was largely restricted to the available $J_{H3'P}$ couplings, derived from the ^{1}H-^{31}P heterocorrelation spectrum. Unfortunately, this coupling constant has the same value for the two spatially allowed *trans* and *gauche*(–) regions and can consequently not discriminate between these conformations. A long range coupling between H2′ and P3 was not observed in the ^{1}H-^{31}P heterocorrelation spectrum, suggesting that the ε angles of all residues reside in the *trans* region.

Determination of glycosidic torsion angles

Analysis of the intra-residue NOE cross-peaks showed that the χ angles of all residues fall in the *anti* domain normally found in B-DNA, except for residue A10. A high intensity of the H8–H1′ cross-peak was observed, together with H8–H2′ and H8–H2″ cross-peaks of the same intensity as the corresponding cross-peaks of other residues. Earlier this observation suggested to us that A10 is in the *syn* orientation (23). A more detailed analysis provides a better definition of this angle, i.e. integration of the NOE volumes of the intra-residue cross-peaks, subsequent distance calculation using the NO2DI program (26) and consideration of the P–χ correlation plots for H8–H1′, H8–H2′ and H8–H2″ proton pairs (29) shows that χ of residue A10 adopts a value between 0 and –60°.

Additional evidence for this orientation of A10 comes from the chemical shift position observed for the H2′ sugar proton resonance of this residue (2.67 p.p.m.). The ring current effect experienced by this proton is very sensitive to the value of the glycosidic torsion angle. As a consequence, χ in the *syn* domain would have resulted in a substantial downfield shift of the H2′ resonance, which is definitely not the case (Wijmenga *et al.*, manuscript in preparation).

Structure calculations

The distance restraints together with the allowed dihedral angle domains and sugar pseudorotation parameters served as input for

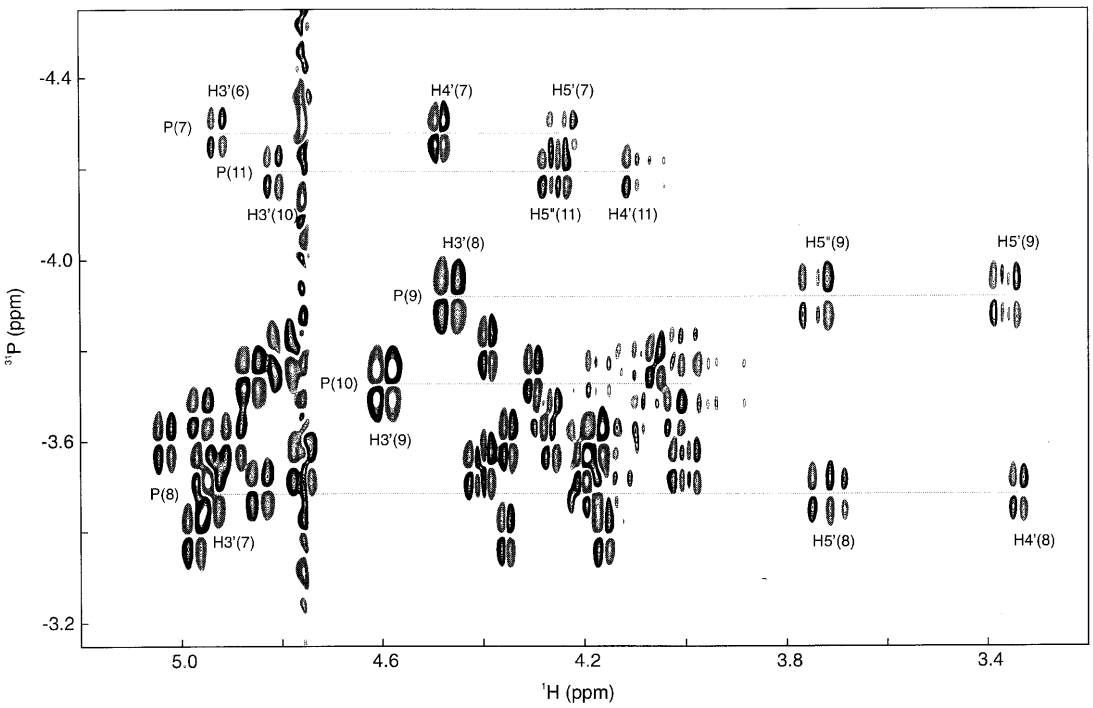

Figure 3. ^{31}P-^{1}H correlation spectrum of d(ATCCTA-GTTA-TAGGAT) at 298 K. Indicated are the resonance positions of the phosphorous atoms of residues C8, C9 and G10. Assignments of the cross-peaks concerning loop residues are indicated.

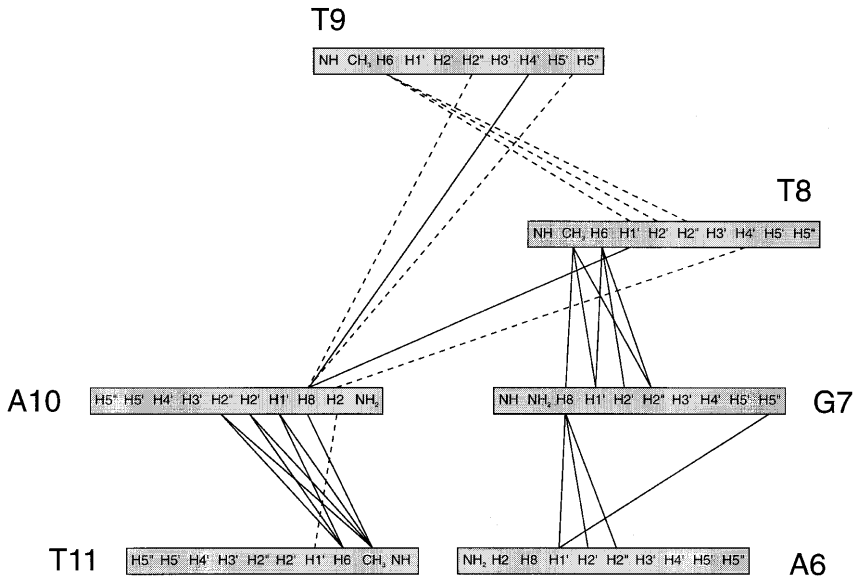

Figure 4. Schematic representation of the inter-residue NOE connectivities in the loop region of the hairpin formed by d(ATCCTA-GTTA-TAGGAT). Strong and weak NOEs are indicated by solid and dashed lines respectively.

the distance geometry program DIANA. These data were supplemented by hydrogen bond restraints for the formation of a G-A base pair, the formation of which was suggested by thermodynamic studies performed earlier (23). Basically, a number of G-A pair configurations can possibly exist (see, for instance, 13,37,38), but in a first consideration only those in which both glycosidic torsion angles reside inside the *anti* domain qualify. We tried all of the possible candidates as starting configurations in the structure calculations but only that with hydrogen bonds between the N3 atom of G7 and the amino group of A10 and between the amino group of G7 and the N7 atom of A10 did not lead to severe violations of the NMR constraints. On the contrary. Although we did not have direct NOE evidence for hydrogen bond formation between the G7 and A10 amino protons, the latter base pair fitted the NMR restraints excellently. It should be mentioned in passing that the imino proton of G7 is

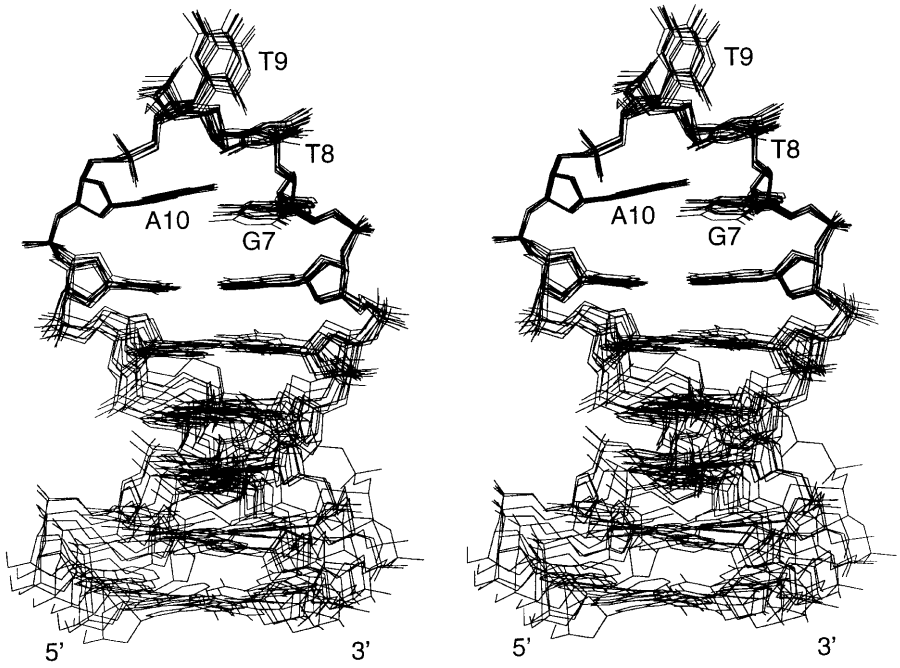

Figure 5. Ensemble of the 10 structures with lowest energies resulting from NOE and dihedral restraints derived for the d(ATCCTA-GTTA-TAGGAT) hairpin. The loop region, i.e. residues 6–11, was aligned before superposition to give the best possible fit.

not involved in base pairing, in accordance with its resonance position and temperature-dependent behaviour (see Fig. 1).

The subsequent structure refinement calculations within X-PLOR mainly affected locally unfavourable conformations. Structures with high energies after these calculations were discarded from further analysis. The procedure resulted in a set of 24 refined structures, of which the average root mean square (r.m.s.) deviation for the loop region was 0.33 Å with respect to the mean structure. Figure 5 shows an overlay of the 10 structures, taken from this set, with lowest energies resulting from NMR-derived distance and dihedral restraints. The set comprises structures with both a β(9) angle of ~110° as well as ~250° (*vide supra*). However, energy minimization of these refined structures, in a force field which included terms for electrostatic and dihedral contributions, revealed that the first situation is energetically more favourable, which indicates that it is more likely to occur.

Conformation of the backbone

The complementary ends of the d(ATCCTA-GTTA-TAGGAT) fragment form the six base pairs of a B-type stem, connected by a four-membered loop, which changes the direction of the phosphate backbone by 180° (Fig. 6). As in all DNA tetraloops studied previously, this change in backbone direction is not spread throughout the entire loop region but takes place within a relatively small portion of the loop, i.e. between the third and fourth loop residues. Up to this point, the backbone essentially follows the general course of the 5′-strand of the underlying helix, although it tends to shift a little in the direction of the helix axis. After the turn the course of the opposite strand is followed.

Examining the turn in the phosphate backbone in greater detail shows that the GTTA loop behaves somewhat unconventionally when compared with other DNA tetraloops. Although the conformational studies performed thus far showed that such a sharp turn can be accomplished by various combinations of dihedral angles, the value for the γ angle of the fourth residue has consequently been found in the *trans* region. However, γ(10) in the GTTA loop (91°) deviates only slightly from the usual helical value. Instead, the turn is essentially accomplished by adjustments of ζ(9) and α(10), which are converted to the *gauche*+ and *trans* regions respectively. Other alterations with respect to standard helical values are observed for ε(9) and β(10), both of which are transposed to the (–)anticlinal domain, as well as for two additional dihedral angles of T9. The value for γ(9) in the *trans* region combined with β(9) (+)anticlinal induces the backbone to deviate gradually from the cylinder wall (see Fig. 6), preparing in this way the actual turn between T9 and A10.

The G-A base pair

The GTTA loop exhibits a compact structure, dominated by stacking interactions. Figure 7a shows that the first base in the loop, belonging to G7, stacks quite extensively on the base of the last stem residue, A6. Compared with the situation encountered for normal helix propagation, the base of G7 is slightly shifted towards the major groove and its sugar moiety lies more inwards with respect to the helix cylinder. The base of residue A10 also stacks on the stem, i.e. on the base of T11, but not at all in a regular fashion. It is drastically shifted towards the minor groove, partly accomplished by the glycosidic torsion angle of this residue being ~–70°. In this way the N6/N7 flank of A10 is in close proximity to the N2/N3 flank of G7, thus enabling the formation of two hydrogen bonds between these residues. The stacking interaction between A10 and T11 is rather poor.

To accommodate the sheared G-A base pair, the value for the helical twist for this base pair is substantially lower than those for the stem, i.e. 13.0° for the G-A base pair versus 36.1° on average for the base pairs in the stem. The small helical twist brings the

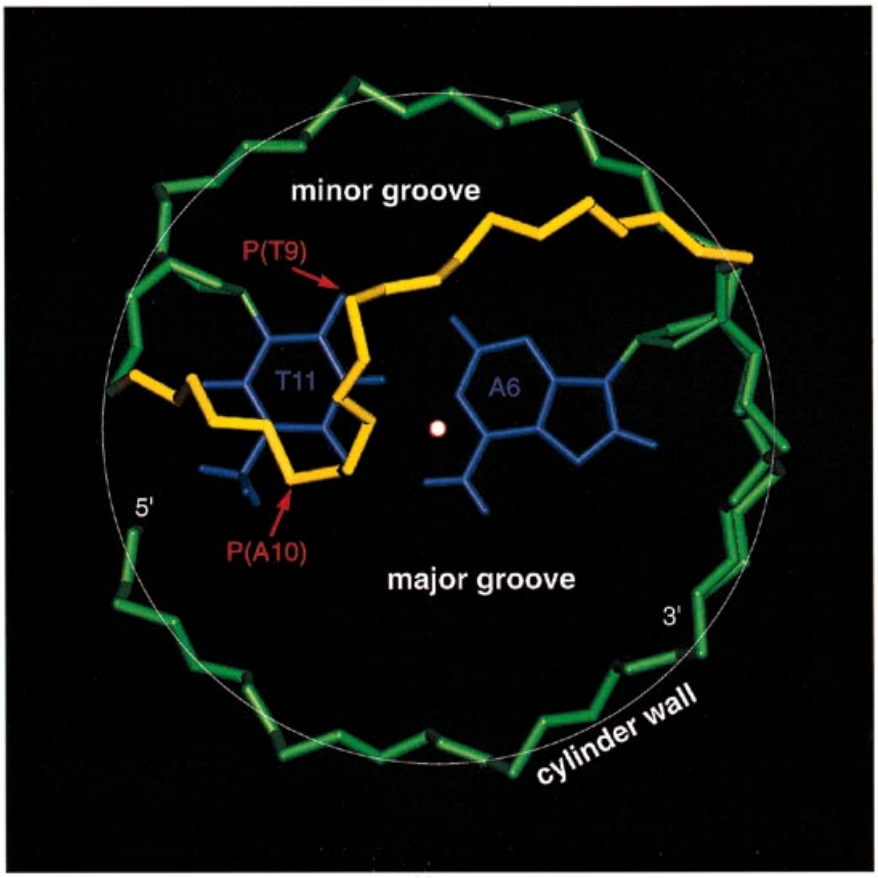

Figure 6. View along the helix axis of the average structure of d(ATCCTA-GTTA-TAGGAT) after energy minimization, indicating the course of the phosphate backbone. The backbone for the residues in the GTTA loop is shown in yellow; the remainder of the backbone is shown in green. The central white dot represents the position of the helix axis.

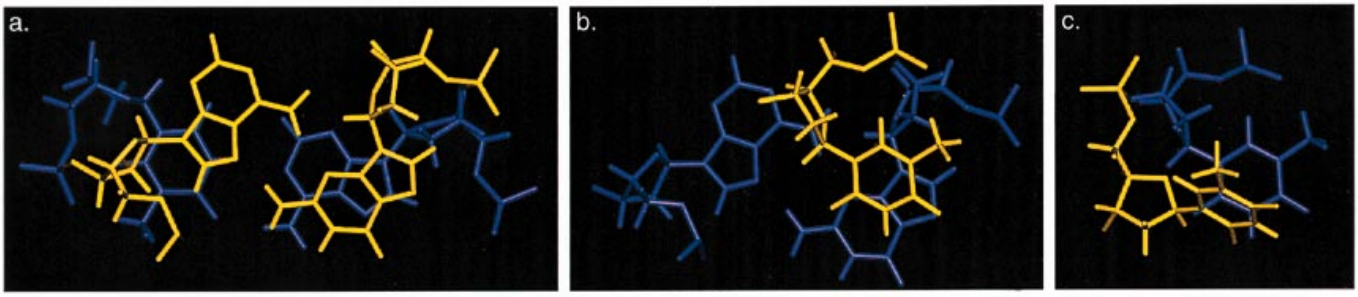

Figure 7. Portions of the average structure of d(ATCCTA-GTTA-TAGGAT) after energy minimization, as viewed from the top, illustrating the relative positions of (**a**) the G7-A10 base pair and the underlying A6·T11 base pair, (**b**) the unpaired T8 residue and the G7-A10 base pair and (**c**) the unpaired T9 and T8 residues. In all instances, the base (pair) lying on top is shown in yellow, the base (pair) lying below is shown in blue.

base of A10 directly over H1′ of T11, resulting in the unusual upfield resonance position of this proton (see Table 3). A similar effect has been observed in various other compounds comprising one or more sheared G-A base pairs (5,7,16,17). Furthermore, we observed values for the tip, inclination and propeller twist for the G-A base pair (14.9, 5.0 and 9.6° respectively) higher than the average values for the stem (0.0, 1.5 and –4.3° respectively).

Type-I loop folding

From a detailed comparison of the accumulated structural data (1) it follows that stable hairpin loops tend to fold into a limited number of well-defined, compact conformations. For tetraloops, three distinct types of folding, designated type-I, type-II and type-III folds (24), have been observed. Type-I loops, observed exclusively for DNA, adopt a conformation with the first three bases at the 5′-end of the loop forming a more or less continuous stack on the 3′-end of the stem. In type-II loops, found in DNA as well as in RNA, the base of the second loop residue from the 5′-end of the loop is turned into or towards the minor groove and the base of the third residue lies over the base pair formed by residues 1 and 4. The third fold, type-III, is only observed in RNA and is described by a continuous stacking of the loop bases on the 5′-end of the stem.

Of the unpaired loop residues in the GTTA loop, T8 stacks upon its 5′-neighbour (Fig. 7b), something which is not entirely true for T9. The latter is directed away from the loop and points into the solvent (Fig. 7c). Nevertheless, the GTTA loop structure established here can be considered as belonging to the type-I category of loop folding. Hence, the results of this study conform to the idea of the existence of a small number of folding patterns for four-membered loops.

For tetraloops containing a Watson–Crick base pair (miniloops) it has been found that the position of the first unpaired base may depend upon its nature. It has been proposed that if this residue is a purine, it tends to stack upon its 5′-neighbour, whereas a pyrimidine residue preferentially folds into the minor groove (39). This rule also holds for tetraloops comprising various other types of base pairs (24,40,41). Therefore, one might have expected that the second residue in the GTTA loop would be positioned in the minor groove. The fact that this is not the case can easily be understood by considering the geometry of the sheared G-A base pair. In this arrangement the adenine base has shifted towards the minor groove (see Fig. 7), which prevents T8 from folding into this groove. Hence, the results presented here suggest that the overall folding of a loop containing a G-A base pair is not determined by the nature of the second base. A similar stacking pattern might be expected therefore for loops with a purine as the second loop residue. The comparable melting temperatures for d(ATCCTA-GTTA-TAGGAT) and d(ATCCTA-GAAA-TAGGAT), i.e. 55 and 53°C respectively (23), as well as the results of Sandusky *et al.* (8), corroborate this.

Interaction of the G-A base pair and the closing base pair of the stem

A number of conformational studies have been performed in which a sheared G-A base pair was found in a four-membered loop. The majority of these investigations concerned loops meeting the consensus sequence GNRA, a motif which is often denoted as unusually stable. However, the actual stability of these loops varies strongly and the question might be raised as to whether all of them really contain a sheared G-A base pair.

Recently, Sandusky *et al.* (8) observed that the stability of GNNA tetraloops strongly depends on the polarity of the C·G base pair lying underneath, a phenomenon also known to occur for other hairpin loops (1). The melting temperature of a loop composed of C-GGCA-G was found to be almost 20°C higher than that of a G-GGCA-C sequence and it was concluded that the presence of a sheared G-A pair in the second case was improbable. The authors explained this observation by predicting very poor stacking interactions for the hypothetical situation of a sheared G-A base pair positioned on top of a G·C base pair, especially between the two consecutive purines (GG). These findings might suggest that the rule formulated by Reid and co-workers, i.e. a sheared G-A base pair configuration is only formed when the preceding base pair is C·G or T·A (19), also applies to G-A base pairs in loops. However, the results presented here show otherwise. We find a well-defined loop structure, including a sheared G-A base pair, on top of an A·T moiety. The stacking interactions between these two base pairs are definitely much better than those predicted for the 5′-GG-3′/3′-CA-5′ unit (8). In particular, the A6–G7 stacking encountered for the A-GTTA-T loop seems to be of the same order as the adenine–guanine interaction in the structure adopted by the 5′-CG-3′/3′-GA-5′ unit. Similar results as observed here were found earlier for an RNA GAGA tetraloop positioned on top of an A·U base pair (7). Unfortunately, we are not aware of thermodynamic or structural data for loops composed of a T(U)-GNNA-A sequence. Still, the structure of the T-GCA-A loop (9) indicates that favourable stacking interactions are possible for such a situation.

In conclusion, a picture emerges that the stability of a loop containing a sheared G-A base pair is relatively insensitive to the polarity of an underlying A·T(U) base pair, contrary to the situation in which this is a C·G base pair. This effect is possibly related to the direction of the dipole moment of the guanine base, which is more or less opposite relative to those of thymine, adenine and cytosine (42). This might result in a more favourable situation for CG, AG and TG stacks, as compared with a GG stack, and therefore might lead to more favourable stacking interactions in 5′-CG-3′/3′-GA-5′, 5′-AG-3′/3′-TA-5′ and 5′-TG-3′/3′-AA-5′ units and less favourable interactions for 5′-GG-3′/3′-CA-5′.

ACKNOWLEDGEMENTS

NMR spectra were recorded at the Dutch National hf-NMR Facility (Nijmegen, The Netherlands), supported by SON. We wish to thank J.Joordens for excellent technical assistance.

REFERENCES

1 Hilbers,C.W., Heus,H.A., van Dongen,M.J.P. and Wijmenga,S.S. (1994) In Eckstein,F. and Lilley,D.M.J. (eds), *Nucleic Acids and Molecular Biology.* Springer-Verlag, Berlin, Germany, Vol. 8, pp. 56–104.
2 Hirao,I., Nishimura,Y., Naroka,T., Watanabe,K., Arata,Y. and Miura,K.-i. (1989) *Nucleic Acids Res.*, **17**, 2223–2231.
3 Hirao,I., Nishimura,Y., Tagawa,Y.-i., Watanabe,K. and Miura,K.-i. (1992) *Nucleic Acids Res.*, **20**, 3891–3896.
4 Hirao,I., Nishimura,Y., Tagawa,Y.-i., Watanabe,K. and Miura,K.-i. (1994) *Nucleic Acids Res.*, **22**, 576–582.
5 Heus,H.A. and Pardi,A. (1991) *Science*, **253**, 191–193.
6 SantaLucia,J.,Jr, Kierzek,R. and Turner,D.H. (1992) *Science*, **256**, 217–219.
7 Orita,M., Nishikawa,F., Shimanaya,T., Taira,K., Endo,Y. and Nishikawa,S. (1993) *Nucleic Acids Res.*, **21**, 5670–5678.
8 Sandusky,P., Wooten,E.W., Kurochkin,A.V., Kavanaugh,Th., Mandecki,W. and Zuiderweg,E. (1995) *Nucleic Acids Res.*, **23**, 4717–4725.
9 Zhu,L., Chou,S.-H., Xu,J. and Reid,B.R. (1995) *Nature Struct. Biol.*, **2**, 1012–1017.
10 Woese,C.R., Winker,S. and Gutell,R.R. (1990) *Proc. Natl. Acad. Sci. USA*, **87**, 8467–8471.
11 Catasti,P., Gupta,G., Garcia,A.E., Ratliff,R., Hong,L., Yau,P., Moyzis,R.K. and Bradbury,E.M. (1994) *Biochemistry*, **33**, 3819–3830.
12 Ferrer,N., Azorín,F., Villasante,A., Gutierrez,C. and Abad,J.P. (1995) *J. Mol. Biol.*, **245**, 8–21.
13 Li,Y., Zon,G. and Wilson,W.D. (1991) *Proc. Natl. Acad. Sci. USA*, **88**, 26–30.
14 Li,Y., Zon,G. and Wilson,W.D. (1991) *Biochemistry*, **30**, 7566–7572.
15 Chou,S.-H., Cheng,J.W. and Reid,B.R. (1992) *J. Mol. Biol.*, **228**, 138–155.
16 Chou,S.-H., Cheng,J.W., Fedoroff,O. and Reid,B.R. (1994) *J. Mol. Biol.*, **241**, 467–479.
17 SantaLucia,J.,Jr and Turner,D.H. (1993) *Biochemistry*, **32**, 12612–12623.
18 Gautheret,D., Konings,D. and Gutell,R.R. (1994) *J. Mol. Biol.*, **242**, 1–8.
19 Cheng,Y.K., Chou,S.-H. and Reid,B.R. (1992) *J. Mol. Biol.*, **228**, 1037–1041.
20 Walter,A.E., Wu,M. and Turner,D.H. (1994) *Biochemistry*, **33**, 11349–11354.
21 Antao,V.P. and Tinoco,I.,Jr (1992) *Nucleic Acids Res.*, **20**, 819–824.
22 Jucker,F.M., Heus,H.A., Ping,F.Y., Moors,E.H.M. and Pardi,A. (1996) *J. Mol. Biol.*, **264**, 968–980.

23 Hilbers,C.W., Blommers,M.J.J., Van de Ven,F.J.M., Van Boom,J.H. and Van der Marel,G.A. (1991) *Nucleosides Nucleotides*, **10**, 61–80.
24 Van Dongen,M.J.P., Wijmenga,S.S., Van der Marel,G.A., Van Boom,J.H. and Hilbers,C.W. (1996) *J. Mol. Biol.*, **263**, 715–729.
25 Van Boom,J.H., Van der Marel,G.A., Westerink,H.P., Van Boeckel,C.A.A., Mellema,J.R., Altona,C., Hilbers,C.W., Haasnoot,C.A.G., de Bruin,S.H. and Berendsen,R.G. (1983) *Cold Spring Harbor Symp. Quant. Biol.*, **47**, 403–409.
26 Van de Ven,F.J.M., Blommers,M.J.J., Schouten,R.E. and Hilbers,C.W. (1991) *J. Magn. Resonance*, **94**, 140–151.
27 Cantor,C.R. and Schimmel,P.R. (1980) *Biophysical Chemistry*, Vol. II. W.H.Freeman and Co., San Francisco, CA.
28 Blommers,M.J.J., Van de Ven,F.J.M., Van der Marel,G.A., Van Boom,J.H. and Hilbers,C.W. (1991) *Eur. J. Biochem.*, **201**, 33–51.
29 Wijmenga,S.S., Mooren,M.M.W. and Hilbers,C.W. (1993) In Roberts,G.C.K. (ed.), *NMR of Macromolecules, A Practical Approach*. Oxford University Press, New York, NY, pp. 217–288.
30 Güntert,P. (1995) *DIANA User's Manual and Instructions, Version 2.8*. Institut für Molekularbiologie und Biophysik, ETH. Zurich, Switzerland.
31 Güntert,P., Braun,W. and Wüthrich,K. (1991) *J. Mol. Biol.*, **217**, 517–530.
32 Saenger,W. (1984) *Principles of Nucleic Acid Structure*, Springer Verlag, New York, NY
33 Wüthrich,K. (1986) *NMR of Proteins and Nucleic Acids*. John Wiley & Sons, New York, NY.
34 Keepers,J.W. and James,Th.L. (1984) *J. Magn. Resonance*, **57**, 404–426.
35 Mooren,M.M.W., Pulleyblank,D.E., Wijmenga,S.S., Van de Ven,F.J.M. and Hilbers,C.W. (1994) *Biochemistry*, **33**, 7315–7325.
36 Brünger,A.T. (1992) *X-PLOR Manual, Version 3.1*. Yale University Press, New Haven, CT.
37 Brown,T., Hunter,W.N., Kneale,G. and Kennard,O. (1986) *Proc. Natl. Acad. Sci. USA*, **83**, 2402–2406.
38 Gao,X.L. and Patel,D.J. (1988) *J. Am. Chem. Soc.*, **110**, 5178–5182.
39 Ippel,J.H. (1993) Loop folding in DNA minihairpins. Conformational studies. Thesis University of Leiden, Leiden, The Netherlands, pp 209–239.
40 Varani,G., Cheong,C. and Tinoco,I.,Jr (1991) *Biochemistry*, **30**, 3280–3289.
41 Allain,F.H.-T. and Varani,G. (1995) *J. Mol. Biol.*, **250**, 333–353.
42 Bloomfield,V.A., Crothers,D.M. and Tinoco,I.,Jr (1974) *Physical Chemistry of Nucleic Acids*. Harper & Row, New York, NY
43 Sklenar,V. and Bax,A. (1987) *J. Magn. Resonance*, **74**, 469–479.
44 Jeener,J., Meier,B.H., Bachmann,P. and Ernst,R.R. (1979) *J. Chem. Phys.*, **71**, 4546–4553.
45 Marion,D. and Wüthrich,K. (1983) *Biochem. Biophys. Res. Commun.*, **113**, 967–974.
46 Rance,M., Sørensen,O.W., Bodenhausen,G., Wagner,G., Ernst,R.R. and Wüthrich,K. (1983) *Biochem. Biophys. Res. Commun.*, **117**, 479–485.
47 Bax,A. and Davis,D.G. (1985) *J. Magn. Resonance B*, **65**, 355–366.
48 Griesinger,C., Otting,G., Wüthrich,K. and Ernst,R.R. (1988) *J. Am. Chem. Soc.*, **110**, 7870–7872.
49 Sklenar,V., Miyashiro,H., Zon,G., Miles,H.T. and Bax,A. (1986) *FEBS Lett.*, **208**, 94–98.